

# Entropic Dynamics<sup>\*</sup>

Ariel Caticha

Department of Physics, University at Albany-SUNY,  
Albany, NY 12222, USA.<sup>†</sup>

## Abstract

I explore the possibility that the laws of physics might be laws of inference rather than laws of nature. What sort of dynamics can one derive from well-established rules of inference? Specifically, I ask: Given relevant information codified in the initial and the final states, what trajectory is the system expected to follow? The answer follows from a principle of inference, the principle of maximum entropy, and not from a principle of physics. The entropic dynamics derived this way exhibits some remarkable formal similarities with other generally covariant theories such as general relativity.

## 1 Introduction

The study of changes in the natural world, dynamics, is divided among several distinct disciplines. Thermodynamics, for example, considers changes between special states, the so-called states of equilibrium, and addresses the question of which final states can be reached from any given initial state. Mechanics studies the changes we call motion, chemistry deals with chemical reactions, quantum mechanics with transitions between quantum states, and the list goes on.

In all of these examples we want to predict or explain the observed changes on the basis of information that is codified in a variety of ways into what we call the states. In some cases the final state can be predicted with certainty, in others the information available is incomplete and we can, at best, only assign probabilities.

The theory of thermodynamics holds a very special place among all these forms of dynamics. With the development of statistical mechanics by Maxwell, Boltzmann, Gibbs and others, and eventually culminating in the work of Jaynes [1], thermodynamics became the first clear example of a fundamental physical theory that could be derived from general principles of probable inference. The entire theory follows from a clear idea of the subject matter, that is, an appropriate choice of which states one is talking about, plus well-known principles of

---

<sup>\*</sup>Presented at MaxEnt 2001, the 21th International Workshop on Bayesian Inference and Maximum Entropy Methods (August 4-9, 2001, Baltimore, MD, USA).

<sup>†</sup>E-mail: Ariel.Caticha@albany.edu

inference [2], namely, consistency, objectivity, universality and honesty. These principles are sufficiently constraining that they lead to a unique set of rules for processing information: these are the rules of probability theory [3] and the method of maximum entropy [1][4].

There are strong indications that a second example of a dynamics that can be deduced from principles of inference is afforded by quantum mechanics [5]. Many features of the theory, traditionally considered as postulates, follow from the correct identification of the subject matter plus general principles of inference. Briefly, the goal of quantum mechanics is not to predict the behavior of microscopic particles, but rather to predict the outcomes of experiments performed with certain idealized setups. Thus, the subject of quantum theory is not just the particles, but rather the experimental setups. The variables that encode the information relevant for prediction are the amplitudes or wave functions assigned to the setups. These ingredients plus a requirement of consistency (namely, that if there are two ways to compute an amplitude, the two results should agree) supplemented by entropic arguments are sufficient to derive most of the standard formalism including Hilbert spaces, a time evolution that is linear and unitary, and the Born probability rule.

If quantum mechanics, deemed by many to be *the* fundamental theory, can be derived in this way, then it is possible, perhaps even likely, that other forms of dynamics might ultimately reflect laws of inference rather than laws of nature. Should this turn out to be the case, then the fundamental equations of change, or motion, or evolution as the case might be, would follow from probabilistic and entropic arguments and the discovery of new dynamical laws would be reduced to the discovery of what is the necessary information for carrying out correct inferences. Unfortunately, this search for the right variables has always been and remains to this day the major stumbling block in the understanding of new phenomena.

The purpose of this paper is to explore this possible connection between the fundamental laws of physics and the theory of probable inference: Can dynamics be derived from inference? Rather than starting with a known dynamical theory and attempting to derive it, I proceed in the opposite direction and ask: What sort of dynamics can one derive from well-established rules of inference?

In section 2 I establish the notation, define the space of states, and briefly review how the introduction of a natural quantitative measure of the change involved in going from one state to another turns the space of states into a metric space [6]. (Such metric structures have been found useful in statistical inference, where the subject is known as Information Geometry [7], and in physics, to study both equilibrium [8] and nonequilibrium thermodynamics [9].)

Typically, once the kinematics appropriate to a certain motion has been selected, one proceeds to define the dynamics by additional postulates. This is precisely the option I want to avoid: in the dynamics developed here there are no such postulates. The equations of motion follow from an assumption about what information is relevant and sufficient to predict the motion.

In a previous paper [6] I tackled a similar problem. There I answered the question:

**Q1:** Given the initial state and that the system evolves to other states, what trajectory is the system expected to follow?

This question implicitly assumes that there is a trajectory and that information about the initial state is sufficient to determine it. The dynamical law follows from the application of a principle of inference, the method of maximum entropy (ME), to the only information available, the initial state and the recognition that motion occurred. Nothing else. The resulting ‘entropic’ dynamics is very simple: the system moves continuously and *irreversibly* along the entropy gradient.

Thus, the honest, correct answer to the inference problem posed by question Q1 has been given, but the equally important question ‘Will the system in fact follow the expected trajectory?’ remained unanswered. Whether the actual trajectory is the expected one depends on whether the information encoded in the initial state happened to be sufficient for prediction. Indeed, for many systems, including those for which the dynamics is *reversible*, more information is needed.

In section 3 we answer the question:

**Q2:** Given the initial and the final states, what trajectory is the system expected to follow?

Again, the question implicitly assumes that there is a trajectory, that in moving from one state to another the system will pass through a continuous set of intermediate states. And again, the equation of motion is obtained from a principle of inference, the principle of maximum entropy, and not from a principle of physics. (For a brief account of the ME method in a form that is convenient for our current purpose see [10].) The resulting ‘entropic’ dynamics also turns out to be simple: the system moves along a geodesic in the space of states. This is simple but not trivial: the geometry of the space of states is curved and possibly quite complicated.

Important features of this entropic dynamics are explored in section 4. We show that there are some remarkable formal similarities with the theory of general relativity (GR). For example, just as in GR there is no reference to an external physical time. The only clock available is provided by the system itself. It turns out that there is a natural choice for an ‘intrinsic’ or ‘proper’ time. It is a derived, statistical time defined and measured by the change itself. Intrinsic time is quantified change. This entropic dynamics can be derived from a Jacobi-type principle of least action, which we explore both in Lagrangian and Hamiltonian form. Just as in GR there is invariance under arbitrary reparametrizations – a form of general covariance – and the entropic dynamics is an example of what is called a constrained dynamics.

## 2 Quantifying change: geometry

In this section we briefly review how to quantify the notion of change (for more details see [6]). The idea is simple: since the larger the change involved in going

from one state to another, the easier it is to distinguish between them, we claim that change can be measured by distinguishability. Next, using the ME method one assigns a probability distribution to each state. This transforms the problem of distinguishing between two states into the problem of distinguishing between the corresponding probability distributions. The solution is well-known: the extent to which one distribution can be distinguished from another is given by the distance between them as measured by the Fisher-Rao information metric [11][12][7]. Thus, change is measured by distinguishability which is measured by distance.

Let the microstates of a physical system be labelled by  $x$ , and let  $m(x)dx$  be the number of microstates in the range  $dx$ . We assume that a state of the system (*i.e.*, a macrostate) is defined by the expected values  $A^\alpha$  of some  $n_A$  appropriately chosen variables  $a^\alpha(x)$  ( $\alpha = 1, 2, \dots, n_A$ ),

$$\langle a^\alpha \rangle = \int dx p(x) a^\alpha(x) = A^\alpha. \quad (1)$$

A crucial assumption is that the selected variables codify all the information relevant to answering the particular questions in which we happen to be interested. This is a point that we have made before but must be emphasized again: there is no systematic procedure to choose the right variables. At present the selection of relevant variables is made on the basis of intuition guided by experiment; it is essentially a matter of trial and error. The variables should include those that can be controlled or observed experimentally, but there are cases where others must also be included. The success of equilibrium thermodynamics, for example, derives from the fact that a few variables are sufficient to describe a static situation, and being few, these variables are easy to identify. In fluid dynamics, on the other hand, the selection is more difficult. One must include many more variables, such as the local densities of particles, momentum, and energy, that are neither controlled nor usually observed.

The states form an  $n_A$ -dimensional manifold with coordinates given by the numerical values  $A^\alpha$ . To each state we can associate a probability distribution  $p(x|A)$ . The distribution that best reflects the prior information contained in  $m(x)$  updated by the information  $A^\alpha$  is obtained by maximizing the entropy

$$S[p : m] = - \int dx p(x) \log \frac{p(x)}{m(x)}. \quad (2)$$

subject to the constraints (1). The result is

$$p(x|A) = \frac{1}{Z} m(x) e^{-\lambda_\alpha a^\alpha(x)}, \quad (3)$$

where the partition function  $Z$  and the Lagrange multipliers  $\lambda_\alpha$  are given by

$$Z(\lambda) = \int dx m(x) e^{-\lambda_\alpha a^\alpha(x)} \quad \text{and} \quad - \frac{\partial \log Z}{\partial \lambda_\alpha} = A^\alpha. \quad (4)$$

Next, we argue that the change involved in going from state  $A$  to the state  $A + dA$  can be measured by the extent to which the two distributions can be

distinguished. As discussed in [7], except for an overall multiplicative constant, the measure of distinguishability we seek is given by the ‘distance’  $d\ell$  between  $p(x|A)$  and  $p(x|A + dA)$ ,

$$d\ell^2 = g_{\alpha\beta} dA^\alpha dA^\beta , \quad (5)$$

where

$$g_{\alpha\beta} = \int dx p(x|A) \frac{\partial \log p(x|A)}{\partial A^\alpha} \frac{\partial \log p(x|A)}{\partial A^\beta} \quad (6)$$

is the Fisher-Rao metric [11][12]. It turns out that this metric is unique: it is the only Riemannian metric that adequately reflects the fact that the states  $A$  are not ‘structureless points’, but happen to be probability distributions.

To summarize: the very act of assigning a probability distribution  $p(x|A)$  to each state  $A$ , automatically provides the space of states with a metric structure.

### 3 Dynamics and intrinsic time

Given the initial and the final states, what trajectory is the system expected to follow? The key to answering this question lies in the implicit assumption that there exists a trajectory or, in other words, that large changes are the result of a continuous succession of very many small changes. Thus, the difficult problem of studying large changes is reduced to the much simpler problem of studying small changes.

Let us therefore focus on small changes and assume that the change in going from the initial state  $A_i$  to the final state  $A_f = A_i + \Delta A$  is small enough that the distance  $\Delta\ell$  between them is given by

$$\Delta\ell^2 = g_{\alpha\beta} \Delta A^\alpha \Delta A^\beta . \quad (7)$$

To find which states are expected to lie on the trajectory between  $A_i$  and  $A_f$  we reason as follows. In going from the initial to the final state the system must pass through a halfway point, that is, a state  $A$  that is equidistant from  $A_i$  and  $A_f$  (see fig.1a). The question is which halfway state should we choose? An answer to this question would clearly determine the trajectory: first find the halfway point, and use it to determine ‘quarter of the way’ points, and so on.

Next we notice that there is nothing special about halfway states. We could equally well have argued that in going from the initial to the final state the system must first traverse a third of the way, that is, it must pass through a state that is twice as distant from  $A_f$  as it is from  $A_i$ . In general, we can assert that the system must pass through intermediate states  $A_\omega$  such that, having already moved a distance  $d\ell$  away from the initial  $A_i$ , there remains a distance  $\omega d\ell$  to be covered to reach the final  $A_f$ . Halfway states have  $\omega = 1$ , ‘third of the way’ states have  $\omega = 2$ , and so on (see fig.1b).

It appears that each different value of  $\omega$  provides a different criterion to select the trajectory. If there is a trajectory and there are several ways to determine it, consistency demands that all these ways should agree: in the end we must verify that the selected trajectory is independent of  $\omega$  or else we have a problem.

Our basic dynamical question Q2 can be rephrased as follows:

Figure 1: The geometry of dynamics: in going from the initial  $A_i$  to the final state  $A_f$  the system must pass through states  $A$  (dashed line) that are equidistant between them (a), and it must also pass through states  $A_\omega$  (dashed circle) that are  $\omega$  times as far from  $A_f$  as they are from  $A_i$  (b).

**Q2'** The system is initially in state  $p(x|A_i)$  and we are given the new information that the system has moved to one of the neighboring states in the family  $p(x|A_\omega)$ . Which  $p(x|A_\omega)$  do we select?

Phrased in this way it is clear that this is precisely the kind of problem to be tackled using the ME method. Recall [10]: The ME method is a method for processing information and change our minds. It allows us to go from an old set of beliefs, described by the prior probability distribution, to a new set of beliefs, described by the posterior distribution, when the information available is just a specification of the family of distributions from which the posterior must be selected [13]. In the more traditional applications of the method this family of posteriors is constrained or defined by the known expected values of some relevant variables, but this is not necessary, the constraints need not be linear functionals. Here the constraints are defined geometrically.

An important question that should arise whenever one contemplates using the ME method is which entropy should one maximize. Since we want to select a distribution  $p(x|A)$  the entropies to be considered must be of the form

$$S[p : q] = - \int dx p(x|A) \log \frac{p(x|A)}{q(x)}. \quad (8)$$

This is the entropy of  $p(x|A)$  relative to the *prior*  $q(x)$ . The interpretation of  $q(x)$  as the prior follows from the logic behind the ME method itself. Recall [10]: In the absence of new information there is no reason to change one's mind. When there are no constraints the selected posterior distribution should coincide with the prior distribution. Since the distribution  $p$  that maximizes  $S[p : q]$  subject to no constraints is  $p \propto q$ , we must set  $q(x)$  equal to the prior.

Coming back to our dynamical problem, suppose we know that the system is initially in state  $p(x|A_i)$  and we are not given the information that the system moved, then we have no reason to suspect that any change has occurred. Therefore the prior  $q(x)$  should be chosen so that the maximization of  $S[p : q]$  subject to no constraints yields the posterior  $p = p(x|A_i)$ . The correct choice is  $q(x) = p(x|A_i)$ .

Now we are ready to tackle the question Q2': the answer is obtained by maximizing the entropy

$$S[A : A_i] = - \int dx p(x|A) \log \frac{p(x|A)}{p(x|A_i)}, \quad (9)$$

subject to the constraint  $A = A_\omega$ . This presents no problems. It is convenient to write  $A_\omega = A_i + dA$  and  $A_f = A_i + \Delta A$  so that  $S[A_\omega : A_i]$  simplifies to

$$S[A_i + dA : A_i] = -\frac{1}{2} g_{\alpha\beta} dA^\alpha dA^\beta, \quad (10)$$

and the distances  $d\ell_i$  and  $d\ell_f$  from  $A_\omega$  to  $A_i$  and  $A_f$  are given by

$$d\ell_i^2 = g_{\alpha\beta} dA^\alpha dA^\beta \quad \text{and} \quad d\ell_f^2 = g_{\alpha\beta} (\Delta A^\alpha - dA^\alpha)(\Delta A^\beta - dA^\beta). \quad (11)$$

Then, to maximize  $S[A_i + dA : A_i]$  under variations of  $dA$  subject to the constraint

$$\omega d\ell_i = d\ell_f, \quad (12)$$

introduce a Lagrange multiplier  $\lambda$ ,

$$\delta \left( -\frac{1}{2} g_{\alpha\beta} dA^\alpha dA^\beta - \lambda (\omega^2 d\ell_i^2 - d\ell_f^2) \right) = 0. \quad (13)$$

We get

$$dA^\alpha = \chi \Delta A^\alpha \quad \text{where} \quad \chi \equiv \frac{1}{1 - \omega^2 + 1/2\lambda}. \quad (14)$$

The multiplier  $\lambda$ , or equivalently the quantity  $\chi$ , is determined substituting back into the constraint (12). From eq.(11) we get  $d\ell_i = \chi \Delta \ell$  and  $d\ell_f = (1 - \chi) \Delta \ell$ , and therefore [14]

$$\chi = \frac{1}{1 + \omega} \quad \text{and} \quad \lambda = \frac{1}{2\omega(1 + \omega)}. \quad (15)$$

Thus, the intermediate state  $A_\omega$  selected by the maximum entropy method is such that

$$d\ell_i + d\ell_f = \Delta \ell. \quad (16)$$

The geometrical interpretation is obvious: the triangle defined by the points  $A_i$ ,  $A_\omega$ , and  $A_f$  (fig.1) degenerates into a straight line. This is sufficient to determine a short segment of the trajectory: all intermediate states lie on the straight line between  $A_i$  and  $A_f$ . The generalization beyond short trajectories is immediate: if any three nearby points along a curve lie on a straight line the curve is a geodesic. Note that this result is independent of the value  $\omega$  so the potential consistency problem we had identified earlier does not arise.

To summarize, the answer to our question Q2 is simple and elegant:

**ED** The expected trajectory is the geodesic that passes through the given initial and final states.

This is the main result of this paper. As promised, in entropic dynamics the motion is predicted on the basis of a ‘principle of inference’, the principle of maximum entropy, and not from a ‘principle of physics’.

The dynamics ED was derived in an unusual way and one should expect some unusual features. Indeed, they become evident as soon as one asks any question involving time. For example, ED determines the vector tangent to the trajectory  $dA^\alpha/d\ell$ , but not the actual ‘velocity’  $dA^\alpha/dt$ . The reason is not hard to find: nowhere in question Q2 nor in any implicit background information is there any reference to an external time  $t$ .

Additional information is required if one is to find a relation between the distance  $\ell$  along the trajectory and the external time  $t$ . In conventional forms of dynamics this information is implicitly supplied by a ‘principle of physics’, by a Hamiltonian which fixes the evolution of a system relative to external clocks. But Q2 makes no mention of any external universe; the only clock available is the system itself, and our problem becomes one of deciding how this clock should be read. We could, for example, choose one of the variables  $A^\alpha$ , say  $A^1$ , as our clock variable and arbitrarily call it intrinsic time. Ultimately it is best to *define intrinsic time so that motion looks simple*.

A very natural definition consists in stipulating that the system moves with unit velocity, then the intrinsic time  $\tau$  is given by the distance  $\ell$  itself,  $d\tau = d\ell$ . *Intrinsic time is quantified change*. A peculiar consequence of this definition is that intervals between events along the trajectory are not a priori known, they are determined only after the equations of motion are solved and the actual trajectory is determined. This reminds us of the theory of General Relativity (GR).

An important feature of GR is the absence of references to an external time. Given initial and final states, in this case the initial and final three-dimensional geometries of space, the proper time interval along any curve between them is only determined after solving the Einstein equations of motion [15]. The absence of an external time has been a serious impediment in understanding the classical theory [16] – because it is not clear which variables represent the true gravitational degrees of freedom – and also in formulating a quantum theory [17] – because of difficulties in defining equal-time commutators.

In the following section we rewrite the entropic dynamics in Lagrangian and



Hamiltonian forms and we point out some further formal similarities between ED and GR.

## 4 Formal developments

The entropic dynamics can be derived from an ‘action’ principle. Since the trajectory is a geodesic, the ‘action’ is the length itself,

$$J[A] = \int_{\eta_i}^{\eta_f} d\eta L(A, \dot{A}), \quad (17)$$

where  $\eta$  is an arbitrary parameter along the trajectory, the Lagrangian is

$$L(A, \dot{A}) = \left( g_{\alpha\beta} \dot{A}^\alpha \dot{A}^\beta \right)^{1/2} \quad \text{and} \quad \dot{A}^\alpha = \frac{dA^\alpha}{d\eta}. \quad (18)$$

The action  $J[A]$  is invariant under reparametrizations  $A(\eta) \rightarrow A(f(\eta))$  provided the end points are left unchanged,  $f(\eta_i) = \eta_i$  and  $f(\eta_f) = \eta_f$ . Indeed, when the transformation is infinitesimal,  $f(\eta) = \eta + \varepsilon(\eta)$ , the corresponding change in the action,

$$\delta J = \left( g_{\alpha\beta} \dot{A}^\alpha \dot{A}^\beta \right)^{1/2} \varepsilon(\eta) \Big|_{\eta_i}^{\eta_f}, \quad (19)$$

vanishes provided  $\varepsilon(\eta_i) = \varepsilon(\eta_f) = 0$ . As pointed out in [18] there is an important distinction between the symmetries of a generally covariant theory such as GR and the internal symmetries of a proper gauge theory. The action of a generally covariant theory is invariant under those reparametrizations that are restricted to map the boundary onto itself; for proper internal gauge transformations there are no such restrictions. Thus ED shares with GR the fact that both are generally covariant theories.

It is instructive to consider the analogous principle of least action for a nonrelativistic particle. The standard Hamilton’s principle requires extremizing the action

$$\int_{t_i}^{t_f} dt \left( \frac{m}{2} \delta_{ab} \frac{dx^a}{dt} \frac{dx^b}{dt} - V(x) \right), \quad (20)$$

where  $t$  is ‘physical’ time, and the interval between initial and final states  $t_f - t_i$  is given. In contrast, Jacobi’s principle of least action for a particle with energy  $E$  moving in a potential  $V(x)$  determines the trajectory by extremizing the action

$$J[x] = \int_{\eta_i}^{\eta_f} d\eta \left( 2m\delta_{ab} \frac{dx^a}{d\eta} \frac{dx^b}{d\eta} \right)^{1/2} (E - V(x))^{1/2}. \quad (21)$$

There is no reference to any time  $t$ , the time interval between initial and final states is not given, and the parameter  $\eta$  is unphysical and arbitrary. To

determine the temporal evolution along the trajectory requires an additional supplementary condition,

$$\frac{m}{2} \delta_{ab} \frac{dx^a}{dt} \frac{dx^b}{dt} + V(x) = E . \quad (22)$$

Thus the ED action, eq.(17), is an action of the Jacobi type. The natural choice for a supplementary condition that defines  $\tau$  and determines the evolution along the trajectory is

$$g_{\alpha\beta} \frac{dA^\alpha}{d\tau} \frac{dA^\beta}{d\tau} = 1 . \quad (23)$$

It is interesting that GR is also described by a Jacobi-type action [19]. To explore this similarity further it is convenient to construct the canonical (*i.e.*, Hamiltonian as opposed to Lagrangian) version of Jacobi's action.

The canonical momenta are given by

$$\pi_\alpha = \frac{\partial L}{\partial \dot{A}^\alpha} = \frac{g_{\alpha\beta} \dot{A}^\beta}{\left(g_{\mu\nu} \dot{A}^\mu \dot{A}^\nu\right)^{1/2}} \quad (24)$$

and have unit magnitude,

$$g^{\alpha\beta} \pi_\alpha \pi_\beta = 1 . \quad (25)$$

The canonical Hamiltonian vanishes identically,

$$H_{\text{can}}(A, \pi) = \dot{A}^\alpha \pi_\alpha - L(A, \dot{A}) \equiv 0 , \quad (26)$$

because the Lagrangian is homogeneous of first degree in the  $\dot{A}$ 's. Physically this is not surprising: the generator of time evolution can be expected to vanish whenever there is no external time with respect to which the system could possibly evolve. We are led to consider the canonical action

$$\int_{\eta_i}^{\eta_f} d\eta \left( \dot{A}^\alpha \pi_\alpha - H_{\text{can}} \right) = \int_{\eta_i}^{\eta_f} d\eta \dot{A}^\alpha \pi_\alpha , \quad (27)$$

but eq.(25) implies that unconstrained variations of the momenta  $\pi_\alpha$  are not allowed. The correct variational principle requires to extremize the action

$$I[A, \pi, N] = \int_{\eta_i}^{\eta_f} d\eta \left[ \dot{A}^\alpha \pi_\alpha - N h(A, \pi) \right] \quad (28)$$

where

$$h(A, \pi) = \frac{1}{2} g^{\alpha\beta} \pi_\alpha \pi_\beta - \frac{1}{2} \quad (29)$$

and  $N(\eta)$  are Lagrange multipliers that enforce the constraint

$$h(A, \pi) = 0 . \quad (30)$$

for each value of  $\eta$ . The overall factor of  $1/2$  in eq.(29) is introduced for later convenience; it amounts to rescaling  $N$ . Variation of  $I[A, \pi, N]$  with respect to  $A$ ,  $\pi$ , and  $N$  yields the equations of motion,

$$\dot{\pi}_\alpha = -N \frac{\partial h}{\partial A^\alpha}, \quad \dot{A}^\alpha = N \frac{\partial h}{\partial \pi_\alpha}, \quad (31)$$

and eq.(30). Naturally there is no equation of motion for  $N$  and it must be determined from the constraint. We obtain,

$$N = \left( g_{\alpha\beta} \dot{A}^\alpha \dot{A}^\beta \right)^{1/2}, \quad (32)$$

which, using the supplementary condition eq.(23), implies

$$d\tau = N d\eta. \quad (33)$$

The analogue of  $N$  in GR is called the lapse function, it gives the increase of ‘intrinsic’ time per unit increase of the unphysical parameter  $\eta$ . In terms of  $\tau$  the equations of motion become

$$\frac{d\pi_\alpha}{d\tau} = -\frac{\partial h}{\partial A^\alpha} \quad \text{and} \quad \frac{dA^\alpha}{d\tau} = \frac{\partial h}{\partial \pi_\alpha}. \quad (34)$$

In reparametrization invariant or generally covariant theories there is no canonical Hamiltonian (it vanishes identically) but there are constraints. It is the constraints that play the role of generators of evolution, of change. Accordingly, the analogue of eq.(30) in GR is called the Hamiltonian constraint.

## 5 Final remarks

I have provided an answer to the question ‘Given the initial and final states, what is the trajectory followed by the system?’ The answer follows from established principles of inference without invoking additional ‘physical’ postulates. The entropic dynamics thus derived turns out to be formally similar to other generally covariant theories: the dynamics is reversible; the trajectories are geodesics; the system supplies its own notion of an intrinsic time; the motion can be derived from a variational principle that turns out to be of the form of Jacobi’s action principle rather than the more familiar principle of Hamilton; and the canonical Hamiltonian formulation is an example of a dynamics driven by constraints.

To conclude one should point out that a reasonable physical theory must satisfy two key requirements: the first is that it must provide us with a set of mathematical models, the second is that the theory must identify real physical systems to which the models might possibly apply. The entropic dynamics we propose in this paper satisfies the first requirement, but so far it fails with respect to the second; it may be a reasonable theory but it is not yet ‘physical’. There are formal similarities with the general theory of relativity and one wonders: Is

this a coincidence? Whether GR will in the end turn out to be an example of ED is at this point no more than a speculation. A more definite answer hinges on the still unsettled problem of identifying those variables that describe the true degrees of freedom of the gravitational field [16][17].

## References

- [1] E. T. Jaynes, Phys. Rev. **106**, 620 and **108**, 171 (1957); *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, Ed. by R. D. Rosenkrantz (Reidel, Dordrecht, 1983).
- [2] As commonly used the meaning of the word ‘inference’ includes both probable inference and deductive inference, the latter being a limiting case of the former.
- [3] R. T. Cox, Am. J. Phys. **14**, 1 (1946); *The Algebra of Probable Inference* (Johns Hopkins, Baltimore, 1961).
- [4] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **IT-26**, 26 (1980); Y. Tikhonchinsky, N. Z. Tishby and R. D. Levine, Phys. Rev. Lett. **52**, 1357 (1984) and Phys. Rev. **A30**, 2638 (1984); J. Skilling, ‘The Axioms of Maximum Entropy’, in *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Ed. by G. J. Erickson and C. R. Smith (Kluwer, Dordrecht, 1988).
- [5] A. Caticha, Phys. Lett. **A244**, 13 (1998); Phys. Rev. **A57**, 1572 (1998); ‘Probability and entropy in quantum theory’, in *Maximum Entropy and Bayesian Methods*, Ed. by W. von der Linden et al. (Kluwer, Dordrecht, 1999); Found. Phys. **30**, 227 (2000).
- [6] A. Caticha, ‘Change, time and information geometry’, in *Bayesian Methods and Maximum Entropy in Science and Engineering*, Ed. by A. Mohammad-Djafari, AIP Conf. Proc. **568**, 72 (2001) (online at [arXiv.org/abs/math-ph/0008018](http://arXiv.org/abs/math-ph/0008018)).
- [7] S. Amari, *Differential-Geometrical Methods in Statistics* (Springer-Verlag, 1985); C. C. Rodríguez, ‘The metrics generated by the Kullback number’, in *Maximum Entropy and Bayesian Methods*, Ed. by J. Skilling (Kluwer, Dordrecht, 1989); ‘Objective Bayesianism and geometry’, in *Maximum Entropy and Bayesian Methods*, Ed. by P. F. Fougère (Kluwer, Dordrecht, 1990); and ‘Bayesian robustness: a new look from geometry’, in *Maximum Entropy and Bayesian Methods*, Ed. by G. R. Heidbreder (Kluwer, Dordrecht, 1996).
- [8] F. Weinhold, J. Chem. Phys. **63**, 2479 (1975); G. Ruppeiner, Phys. Rev. A **20**, 1608 (1979) and **27**, 1116 (1983); L. Diósi and B. Lukács. Phys. Rev. A **31**, 3415 (1985) and Phys. Lett. **112A**, 13 (1985); R. S. Ingarden, Tensor, N. S. **30**, 201 (1976).

- [9] R. Balian, Y. Alhassid and H. Reinhardt, Phys. Rep. **131**, 1 (1986); R. Balian, Am J. Phys. **67**, 1078 (1999); R. F. Streater, Rep. Math. Phys. **38**, 419 (1996), and Contemp. Math. **203**, 117 (1997).
- [10] A. Caticha, ‘Maximum entropy, fluctuations and priors’, in *Bayesian Methods and Maximum Entropy in Science and Engineering*, Ed. by A. Mohammad-Djafari, AIP Conf. Proc. **568**, 94 (2001) (online at [arXiv.org/abs/math-ph/0008017](http://arXiv.org/abs/math-ph/0008017)).
- [11] R. A. Fisher, Proc. Cambridge Philos. Soc. **122**, 700 (1925).
- [12] C. R. Rao, Bull. Calcutta Math. Soc. **37**, 81 (1945).
- [13] The terms ‘prior’ and ‘posterior’ are normally used to refer to the distributions appearing in Bayes’ theorem; we retain the same terminology when using the ME method because both Bayes and ME are concerned with the similar goal of processing new information to upgrade from a prior to a posterior. The difference lies in the nature of the information involved: for Bayes’ theorem the information is in the form of data, for the ME method it is a constraint on the family of allowed posteriors. In many important applications of the ME method the prior happens to be uniform (*e.g.*, the ‘postulate’ of equal a priori probabilities in statistical mechanics) and it is easy to overlook its presence and the important role it plays.
- [14] There is a second solution to eq.(13),  $\chi = (1 - \omega)^{-1}$ . It corresponds to minimizing rather than maximizing  $S$ .
- [15] R. F. Baierlein , D. H. Sharp and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).
- [16] J. W. York, ‘Kinematics and Dynamics of General Relativity’, in *Sources of Gravitational Radiation*, Ed. by L. Smarr (Cambridge Univ. Press, 1979); A. Ashtekar, ‘Quantum Mechanics of Geometry’ online as gr-qc/9901023; C. Rovelli, ‘Strings, loops and the others: a critical survey on the present approaches to quantum gravity’, in *Gravitation and Relativity: At the Turn of the Millenium*, Ed. by N. Dadhich and J. Narlikar (Poona Univ. Press, 1998); J. Barbour, Class. Quant. Grav. **11**, 2853 and 2875 (1994).
- [17] K. Kuchar, ‘Time and interpretations of quantum gravity’, in *Proceedings of the 4th Canadian Conference on General Relativity and Astrophysics*, Ed. by G. Kunsatter, D. Vincent, and J. Williams (World Scientific, Singapore, 1992); C. Isham ‘Canonical quantum gravity and the problem of time’, in *Integrable Systems, Quantum Groups and Quantum Field Theories*, Ed. by L. A. Ibort and M. A. Rodriguez (World Scientific, Singapore, 1993).
- [18] C. Teitelboim, Phys. Rev. **D25**, 3159 (1982).
- [19] J. D. Brown and J. W. York, Phys. Rev. **D40**, 3312 (1989); J. Hartle, Phys. Rev. **D38**, 2985 (1988).